

DOKTORI (PhD) ÉRTEKEZÉS TÉZISEI

**PROCESSES AND OUTCOMES IN L2 ENGLISH WRITTEN PERFORMANCE
ASSESSMENT: RATERS' DECISION-MAKING PROCESSES AND AWARDED
SCORES IN RATING HUNGARIAN EFL LEARNERS' COMPOSITIONS**

**AZ ANGOL MINT IDEGEN NYELV ÍRÁSKÉSZSÉGÉNEK MÉRÉSE: AZ
ÉRTÉKELŐK DÖNTÉSEI ÉS AZ ODAÍTÉLT PONTSZÁMOK MAGYAR
TANULÓK ANGOL NYELVŰ FOGALMAZÁSAINAK ÉRTÉKELÉSEKOR**

BUKTA KATALIN

Témavezető:

Nikolov Marianne, DSc

**Pécsi Tudományegyetem
Bölcsészettudományi Kar
Nyelvtudományi Doktori Iskola
Alkalmazott Nyelvészeti Doktori Program**

2007

1. Az értekezés témája, kutatási feladata és szerkezete

1.1. A témaválasztás indoklása

Magyarországon az utóbbi évtizedekben jelentősen megnőtt az igény az idegen nyelvek ismeretére, ezért a különböző nyelvek tanulása és ezzel együtt a nyelvismeret mérése előtérbe került. A folyamat jelentősen felgyorsult az 1990-es években, amikor az orosz nyelv kötelező oktatása megszűnt és más modern nyelveket, főleg angolt és németet kezdtek tanítani (Medgyes, 1992). A megnövekedett igények kielégítésére nem állt rendelkezésre megfelelő számú nyelvtanár, mégis a változások hatással voltak az idegen nyelvek tanítására, új módszereket vezettek be az eredményesebb nyelvtanulás érdekében (Imre, 1995). Az idegen nyelvek tanulásával szorosan összefügg a nyelvvizsga, hiszen az elsajátított nyelvtudás mérése is szükséges. Az ezen a területen sokáig egyeduralkodó magyarországi vizsgarendszer mellett megjelentek a nemzetközi vizsgák (Katona, 1998). A nyelvi tudásszint mérésében bekövetkezett változások további jelentős állomása az idegen nyelvi érettségi vizsga megújítása, amelyet az Európai Referenciakeret (CEFR, 2001) ajánlásai figyelembevételével terveztek. Az új érettségi vizsgát 2005-ben vezették be, az előkészületek azonban az 1990-es évek végére vezethetők vissza, amikor egy nagyszabású projekt (Érettségi Reform Projekt) keretében egy szakértőkből álló csoport kidolgozta az új érettségi vizsga modelljét (Fekete, Major, & Nikolov, 1999). Jelen kutatás témájának kiválasztásában jelentős szerepet játszottak az új érettségi vizsgamodell íráskészséget mérő feladatainak tervezése során szerzett ismeretek és tapasztalatok.

Az íráskészség kialakulása szorosan kapcsolódik az oktatáshoz, ez a készség mind az anyanyelven mind az idegen nyelven szervezett oktatási keretekhez kötődik. Az idegen nyelvi íráskészség mérése az anyanyelvi íráskészség méréséből fejlődött ki (Weigle, 2002). Magyarországon az anyanyelvi és az idegen nyelvi íráskészség mérésének kutatása szerény mértékű (Kiszely, 2003). Angol fogalmazások értékelésével, valamint az értékelők képzésével és az értékelés természetével eddig kevesen foglalkoztak (Kontráné, 2003). Jelen dolgozat témaválasztásával a hasonló hazai kutatásokhoz szeretne hozzájárulni, egyrészt az angol fogalmazások értékelési folyamatainak mélyebb feltárásával, másrészt a kutatás módszertanával, melyet idegen nyelvi tudásszint mérésére hazánkban nem alkalmaztak eddig széles körben.

A fogalmazások értékelőinek értékelési folyamatait gondolkodásuk megfigyelésével lehet vizsgálni. A gondolkodás megfigyelésének egyik módja az úgy nevezett verbális protokoll módszere, amely során a gondolatokat hangosan kimondják és rögzítik. Az így

nyert adatokat a kutatás kérdéseinek figyelembe vételével feldolgozzák és elemzik (Ericsson & Simon, 1993). A hangosan gondolkodtatás módszerével gyűjtött adatok segítségével nyomon követhetők a fogalmazások értékelőinek döntései és a pontszámokon kívül döntéshozatalukba is betekintést nyerhetünk.

Jelen értekezés fő célkitűzése, hogy információt gyűjtsön az íráskészséget értékelők gondolkodásáról és megvizsgálja a különböző értékelési folyamatok jellemzőit. Ezen kívül az értekezés azt is célul tűzte ki, hogy feltárja milyen mértékben figyelnek az értékelők a fogalmazásokra, az értékelő skálára és az értékelési kritériumokra értékelésük folyamán. A kutatás adatai két fő forrásból származnak: egyrészt az értékelők pontjaiból, másrészt az értékelők verbális protokolljaiból. Ez utóbbihoz a dolgozat kilencedik fejezetében ismertetett előtanulmány eredményei nyújtottak jelentős segítséget.

1.2. Kutatási feladatok

A kutatás a fogalmazások értékelőinek megfigyelésére irányul és a következő kutatási kérdésekre kíván választ adni:

1. Mi jellemzi a kompetens és a profi értékelők értékelési folyamatait?
2. Milyen kritériumokra figyelnek a kompetens és a profi értékelők?
3. Hogyan értelmezik a kompetens és a profi értékelők a négy értékelési (a feladat teljesítése, szókincs gazdagsága, nyelvhelyesség és a szöveg megszerkesztettsége) szempontot?
4. Hogyan értelmezik a kompetens és a profi értékelők az írásokat?
5. Hogyan látják az íráskészség értékelését, mint feladatot az értékelők?

1.3. Az értekezés szerkezete

Az értekezés két fő részből áll: az első rész az értekezés első nyolc fejezetében betekintést kíván nyújtani az íráskészség kutatásával kapcsolatos nemzetközi és hazai szakirodalomba. A disszertáció második fő részében bemutatott saját kutatás szempontjából releváns szakirodalom tárgyalása az L1 és L2 íráskészség természetének és modelljeinek bemutatásával és a kommunikatív kompetencia jellemzőinek ismertetésével kezdődik (1. fejezet). Ezen belül áttekintem az anyanyelvi és idegen nyelvi íráskészség hasonlóságait és eltéréseit is. A nyelvtudás mérésének jellemzése után (2. fejezet) az íráskészség mérésének módjait részletezem. Az íráskészség mérésének tárgyalását a performancia jellegű tesztelés (McNamara, 1996) ismertetése követi (3. fejezet). Az értékelési skálák és az értékelés legismertebb modelljei állnak a 4. fejezet középpontjában

(Cumming, Kantor, & Powers, 2002; Lumley, 2000; Milanovic, Saville, & Shuhong, 1996; Wolfe, 1997). A különböző modellek az értékelési folyamatok megfigyelhetőségéről tanúskodnak, az eddigi kutatások többféle résztvevővel, sokféle mérőeszközzel és írással igyekeztek feltárni az értékelők viselkedését.

A magyarországi idegennyelv-oktatás és nyelvtudásmérés hagyományainak tárgyalása után (5. fejezet) az anyanyelvi és idegen nyelvi íráskészség magyarországi kutatásának tárgyalása következik (6. fejezet). Külön fejezet foglalkozik az új érettségi vizsgamodell kidolgozásával, amely az idegen nyelvi tudásszint mérésének jó gyakorlatát illusztrálja (7. fejezet).

Az adatgyűjtést az 1.1.-ben említett verbális protokoll módszerrel végeztem, amely kutatási módszer bemutatása szerepel az értekezés első fő részének utolsó fejezetében (8. fejezet). Az értékelési folyamatok gondolkodási műveletek sorozatai és a különböző kognitív műveletek megfigyelésének egyik módja az úgy nevezett verbális protokoll (Gass & Mackey, 2007). Az adatgyűjtés során hangszalagra rögzített szövegek átíratát a kutatás jellegének megfelelően feldolgozzuk és ezáltal a gondolkodási folyamatok elemezhetővé válnak.

Az értekezés első fő részében az irodalmi áttekintés a kutatás elméleti keretét foglalja össze, és megfelelő háttérrel biztosít a kutatásnak, amely a disszertáció második fő része.

2. A kutatás ismertetése

Az értekezés Kilencedik fejezetében egy kísérleti előtanulmány (Pilot study) keretében a verbális protokoll módszerét próbáltam ki, amellyel az értékelési folyamatok megfigyelhetők. Az előtanulmány eredményei azt bizonyították, hogy a fogalmazásokat az értékelők az analitikus értékelő skála szerint bírálták el, továbbá minden értékelő kialakítja az értékelési szempontok saját rendszerét, amit azután alkalmaz az írárok elbírálásánál. A vizsgálat tapasztalatai hozzájárultak az értekezésben alkalmazott kutatás felépítéséhez: alkalmam nyílt a verbális protokollal gyűjtött adatok feldolgozásának alapelveit kipróbálni.

A disszertáció fő kutatási kérdéseinek megválaszolásához kétféle adatot gyűjtöttem:

- (1) az írárok értékelésekor született pontszámokat és
- (2) az értékelők verbális protokolljait.

A kutatásban 37 angol szakos tanárjelölt vett részt, akik egyetemi tanárképzésük utolsó fázisában egy választható féléves kurzusra iratkoztak be, amely az idegen nyelvi

tudásszint mérésével és értékelésével foglalkozik. A hallgatók tudomásul vették, hogy a szemináriumi anyagot az íráskészség mérésével kapcsolatos feladattal egészítettem ki és vállalták a kutatásban való részvételt. A kutatásban résztvevő hallgatókat a továbbiakban értékelőknek nevezem.

Az adatgyűjtéshez tíz fogalmazást választottam ki, amelyet angol középszintű érettségi vizsgára készülő tanulók írtak az érettségi vizsga követelményeinek megfelelő íráskészséget mérő feladat alapján (Bukta, Gróf, & Sulyok, 2005, p. 23). A fogalmazásokat az értékelők egy hatszintű analitikus értékelő skála segítségével négy szempontból (feladat teljesítése, szókincs gazdagsága, nyelvhelyesség és szerkezeti felépítés) értékelték, minden pontszámhoz szintenként és szempontonként kritériumok tartoztak, kivéve két szintet (3 és 5 pont), az úgy nevezett „üres” szinteket, ami az árnyaltabb értékelést tette lehetővé. Az értékelést hangosan végezték és gondolataikat magnószalagra rögzítették. A hanganyag átírása hozzátartozott a feladathoz, így minden értékelő a saját értékelését foglalta írásba és adta be számítógéplemezen.

Az értékelési feladatot az értékelők képzése előzte meg, ami szerves részét képezi a nyelvi performancia mérésének (Alderson, Clapham, & Wall, 1996). Az egyeztető tréning során az értékelők megismerkedtek az idegen nyelvi tudásszint mérésének legfontosabb elméleti és gyakorlati kérdéseivel, az íráskészség mérésével, valamint felkészültek a kutatásban szereplő fogalmazások értékelésére és az adatgyűjtés módjára. Az adatgyűjtés az értékelési feladat során történt, az értékelők az egyeztető tréning végén szóbeli magyarázat után megkapták az „értékelési csomagot”. Az értékelési csomag tartalmazta az értékelési feladat szóban elhangzott utasításait írásban, a tíz írást, az értékelő skálát, a pontozó lapot, egy magnókazettát és egy számítógéplemezt a hanganyag átírásához, valamint egy véleménykérő lapot.

A verbális protokoll adatait korábbi kutatások (Lumley, 2002) és az előtanulmány adatfeldolgozásának módszerei szerint megtervezett lépésekben végeztem. Miután összevettem a hanganyagokat a szövegátiratokkal, a verbális protokollokat olyan szövegegységekre (text unit) bontottam, amelyek információval szolgáltak a kutatási kérdések megválaszolásához. A szövegegységeket rendszerezni kellett a feldolgozáshoz, ehhez az előtanulmány kódolási módszerét alkalmazva egy új kódolási szisztémát alakítottam ki. A kódolási rendszer a kutatási kérdések és a négy értékelési szempont (feladat teljesítése, szókincs gazdagsága, nyelvhelyesség és szerkezeti felépítés) és az értékelők fókuszának figyelembe vételével alakult ki. A kódolás megbízhatóságát ellenpróbával ellenőriztem.

Az íráskészség értékelésében döntő szerepet játszanak az értékelők tulajdonságai, amely mérési hibához vezethet egyeztető tréning ellenére is. Az értékelők tulajdonságainak és magatartásának megfigyelésében korábbi kutatások az értékelőket többféle szempontból csoportosították (pl. Wolfe, 1997). Jelen kutatásban az értékelők azonos előképzettségűek, mindannyian részt vettek az egyeztető tréningen és egyikük sem rendelkezett tanítási és értékelési tapasztalattal, így mindannyian kezdő értékelőknek számítanak. Csoportosításukat kompetens és profi értékelőkre a referenciapontok (benchmark) alapján végeztem. A referenciapontokat a többéves mérési és értékelési tapasztalatom alapján alakítottam ki. A 37 értékelő a feladat teljesítésére, a szókincs gazdagságára, a nyelvhelyességre és a szerkezeti felépítésre adott pontszámait valamint az összpontszámokat hasonlítottam össze Pearson korreláció segítségével. A kompetens csoportba az a 22 értékelő került, akiknek pontszámai három vagy kevesebb alkalommal erős korrelációt mutattak. A profi értékelők csoportjába az a 15 értékelő került, akiknek négy vagy öt pontszáma mutatott erős korrelációt a referenciapontokkal. A kutatás középpontjában a kompetens és a profi értékelők gondolkodási folyamatainak összehasonlítása áll, így nyílt lehetőség az íráskészség értékelési folyamatainak megfigyelésére.

3. A kutatás eredményei

Az első kutatási kérdés az értékelők értékelési folyamatainak jellemzőivel foglalkozik. Az eredmények azt mutatták, hogy a kompetens és a profi értékelők folyamatai sok hasonlóságot mutatnak, ugyanakkor számos eltérés is megfigyelhető. Az értékelők teljesítménye nem különbözik nemek szerint, a nemek aránya a két csoportban azonos volt. Az értékelés nyelvére vonatkozóan az értékelők nem kaptak utasítást, tehát használhatták az anyanyelvüket és a célnyelvet, illetve a kettő keverékét is. Megfigyeléseim szerint a profi értékelők jelentős része a két nyelv keverékét használták, ami arra enged következtetni, hogy az értékelők jobban tudnak a feladatra összpontosítani, ha nem kell a használt nyelvre odafigyelni. A verbális protokollok hossza igen változatos: mind a kompetens, mind a profi értékelők protokolljai széles skálán mozogtak a szószámot illetően, azonban megfigyelhető volt, hogy a profi értékelők a tíz írásra egyenlő mértékben figyeltek. Az írásokat minden értékelő azonos sorrendben, egytől tízig számozva kapta kézhez, ettől a sorrendtől összesen öt értékelő tért el (14%), egyikük sem

adott magyarázatot a sorrend megváltoztatására. Mivel a kutatás az írások sorrendjére nem tért ki, ezt a változtatást nem vettem figyelembe.

Az írásokat az értékelők négy szempont szerint bírálták el, abban a sorrendben, ahogy a szempontok a skálán megjelennek: a feladat teljesítése, a szókincs gazdagsága, nyelvhelyesség és a szöveg szerkesztettsége. Előfordult azonban, főleg a kompetens értékelőknél, hogy egyik szempontról a másikra váltottak, majd visszatértek, illetve egy-két esetben kihagyták valamelyik szempontot. A korábbi kutatásokban (pl. Milanovic, et al., 1996) említett úgy nevezett értékelés előtti szakasz meglétét, amely során az értékelők felszíni jellemzőket sorolnak fel, eredményeim alátámasztották, a profi értékelőkre volt leginkább jellemző, hogy a külalakra, illetve a tartalomra tettek megjegyzést a nélkül, hogy bármelyik szempontra utaltak volna.

A négy értékelési szempont követése különböző értékelési mintákat eredményezett, az értékeléskor használt stratégiák sorrendje változatos volt és nem csak egyénenként, de sok esetben írásonként is más-más volt. Gyakran az értékelő a szempont azonosítása után egy saját kritériumot nevezett meg, majd felolvasott egy példát és kiválasztotta a pontszámot. Másik értékelési minta szerint az azonosítás után az értékelő felolvasott a skálából (példát idézett), saját szempontot említett és kiválasztotta a pontot. Számos alkalommal az azonosítás után a pontszám megnevezése következett (esetenként példa), majd egy saját kritérium. Ugyanakkor az értékelésre ciklikusság is jellemző volt, az értékelők gyakran visszatértek egy-egy stratégiához. Ezen kívül, a feladat teljesítése szempont más mintát eredményezett: az értékelők, különösen a kompetens csoportban sokkal többször fordultak az íráshoz és olvastak belőle, mint a profi értékelők és sokszor a fogalmazás utasításában meghatározott tartalmi pontokat egyesével hasonlították az íráshoz. Összességében megállapítottam, hogy a profi értékelők hasonló mintákat alkalmaztak és kevesebb váltást figyeltem meg az értékelési folyamataikban.

Az értékelési kritériumok vizsgálata állt a következő kutatási kérdés középpontjában. A megfigyelt stratégiákat négy csoportba osztottam a figyelem iránya szerint: szervezési, értékelési, olvasással kapcsolatos és saját szempontot tükröző stratégiákra. Az értékelők a legtöbb figyelmet az értékelési stratégiákra fordították, ezt az olvasással kapcsolatos stratégiák követték. A legkevesebbet a szervezéssel és a saját kritériumok szerinti értékeléssel foglalkoztak.

A szervezési stratégiák közé a dolgozat sorszámanak és az értékelési szempontnak a megnevezése tartozott, illetve azok a megjegyzések, amelyek az értékelési folyamatra utaltak. Eredményeim azt mutatták, hogy az értékelők a dolgozatot és a kritériumot kevés

kivétellel mindig azonosították, valamint ritkán tettek megjegyzést az értékelési folyamatra. Ez arra enged következtetni, hogy az értékelési folyamat nem okozott gondot és a szempontokat mindannyian elsajátították.

Az értékeléssel kapcsolatos stratégiákat mind a négy értékelési szempontnál tizenegy kategóriába soroltam. Mind a kompetens, mind a profi értékelők hasonló mennyiségű stratégiát alkalmaztak. A négy értékelési szempont közül a kompetens és a profi értékelők egyaránt a legtöbb figyelmet a nyelvtan értékelésére fordították. Ezt korábbi hasonló kutatások is megerősítették: az íráskészség értékelésekor a nyelvhelyességi szempont kerül előtérbe (Cumming, et al., 2002; Lumley, 2000; Kontráné, 2003). Bár az előtanulmányban azt figyeltem meg, hogy az értékelők a legtöbbet a feladat teljesítésének értékelésével foglalkoztak és nem a nyelvhelyességgel, ami azzal magyarázható, hogy az előtanulmányban szereplő szövegek igen rövidek voltak és az értékelők a kommunikatív cél teljesítésére összpontosítottak. A legkevesebb figyelmet a kompetens értékelők a szókincs elbírálására fordították, míg a profi értékelők a feladat teljesítésének értékelésével foglalkoztak legkevésbé.

Az olvasással kapcsolatos stratégiák közé soroltam az értékelő skála kritériumainak felolvasását, a szöveg részleteinek olvasását, a szöveg tartalmának összefoglalását, egyes szavak idézését a szövegből, illetve a feladat utasításának olvasását. A kompetens és a profi értékelők olvasási stratégiái különböző mintákat matattak. Az írásokat a kompetens értékelők sokkal többször olvasták a feladat teljesítésének bírálásakor mint a profi értékelők. Hasonló, de kisebb mértékű különbséget figyeltem meg a nyelvhelyesség értékelésekor: a kompetens értékelők többet olvastak az írásokból, mint a profik. A szókincs pontozásakor viszont a profi értékelők olvastak többet, ez arra enged következtetni, hogy a szókincs elbírálása sokszor okozott nehézséget és többször kellett a szöveget olvasni a megfelelő pontszámok kialakításához. A szöveg megszerkesztettsége szempont megítéléséhez olvastak a legkevesebbet az értékelők. Ezt az értékelési szempontot értékelték utoljára és addigra valószínűleg már ismerték a szöveget annyira, hogy nem kellett sokszor olvasni.

Az értékelők sokszor saját szempontokat használtak az írások elbírálásához, a kompetens és a profi értékelők közötti különbség nem volt nagy, bár a profik többször tettek saját megjegyzést és sokszor utaltak felszíni tulajdonságokra, mint például a szövegek hosszára és olvashatóságára, valamint számos alkalommal utaltak a szöveg írójának nyelvtudására és szándékaira.

A harmadik kutatási kérdés célja megvizsgálni, hogy az értékelők hogyan értelmezik a négy értékelési szempontot. A négy értékelési szempontot (a feladat teljesítése, a szókincs gazdagsága, nyelvhelyesség és a szöveg megszerkesztettsége) az értékelési skála tartalmazta, amit az értékelők döntő többsége ebben a sorrendben bírált el. A hatpontos analitikus értékelési skála négy szinten (1, 2, 4 és 6) tartalmazott két-két kritériumot mind a négy szemponthoz. A 0 szint kritériuma csak a feladat teljesítésénél szerepelt, amennyiben a tanuló nem írt semmit, vagy írása nem felelt meg a feladatnak, a többi szempontot nem kell értékelni. A harmadik és az ötödik szint az úgy nevezett „üres” szint volt és az árnyaltabb értékelést tett lehetővé. Az értékelési folyamatok során megfigyelt értékelési stratégiákat mind a négy szempontból tizenegy azonos kategóriába soroltam, míg az olvasással kapcsolatos stratégiákat öt kategória szerint osztályoztam. Ez a csoportosítás lehetővé tette a négy értékelési szempont megfigyelését és a kompetens és profi értékelők stratégiáinak összehasonlítását. Összességében megfigyeltem, hogy minden értékelő kellő alapossggal bírálta el a tíz írást a négy értékelési szempont szerint, azonban különbségeket is megfigyeltem a két értékelő csoport viselkedésében.

Az első értékelési szempont a feladat teljesítésének elbírálása volt: az egyik szempont szerint a kommunikatív cél teljesítését, a másik szerint a feladatban előírt tartalmi pontok tárgyalását kellett elbírálni. A legnagyobb különbség a kompetens és a profi értékelők között az értékelési és az olvasási stratégiák arányában mutatkozott. A kompetens értékelők lényegesen több olvasással kapcsolatos stratégiát alkalmaztak a feladat teljesítésének megítéléséhez. Az értékelő skála két kritériumára nem fordítottak egyenlő figyelmet: sokkal jobban foglalkoztatta az értékelőket az írás tartalmi pontjainak megléte, mint a kommunikatív cél megvalósítása. Az értékelés során sokszor olvasták fel a skála kritériumait, főleg a kompetens értékelők. A feladat teljesítésének kritériumát a profi értékelők sokszor saját szavaikkal értékelték. A szempont értékelésekor az értékelők ritkán tettek saját megjegyzést: előfordult, hogy utaltak az író kreativitására vagy az írás változatosságára. Gyakran előfordult a feladat teljesítésének értékelésekor, hogy az értékelők másik szempontra váltottak és megjegyzést tettek a szókincs változatosságára vagy nyelvhelyességre, illetve az írás szerkezetére. A feladat értékelésekor fordult elő legtöbbször, hogy az értékelők hiányzó részletekre utaltak, illetve a fogalmazást olvasták fel és hasonlították az íráshoz. Eredményeim azt is kimutatták, hogy a kompetens értékelők az értékelési folyamat elején többször váltottak egyik szempontról a másikra, ami a folyamat végére csökkent és ez arra enged következtetni, hogy az értékelési kritériumok beépítéséhez időre volt szükségük. Azt is megfigyeltem, hogy a kompetens

értékelők olvasással kapcsolatos stratégiái is csökkenő tendenciát mutattak az értékelés folyamán, ami tulajdonítható annak, hogy a pontszám odaítélése kevesebb olvasást igényelt egy idő után.

A szókincs elbírálásakor is találtam hasonlóságokat és eltéréseket a két értékelő csoport folyamatainak megfigyelésekor. Az értékelő skála kritériumaira (a szókincs változatossága és a szavak megfelelő használata) hasonló arányban figyeltek, bár a profi értékelők valamivel több értékelő stratégiát alkalmaztak, mint a kompetens értékelők. A profi értékelők sokszor fogalmazták meg bírálatukat saját szavakkal az értékelő skála kritériumai helyett, többször, mint a kompetens értékelők. A szókincs értékelésekor megfigyeltem azt is, hogy az értékelők sokszor nem tudtak különbséget tenni a nyelvhelyesség és a szókincs értékelése között, a szempontokat váltogatták és sokszor tettek megjegyzést a szavak helyesírására. Az olvasási stratégiák használatában több hasonlóságot is megfigyeltem: igen ritkán foglalták össze az írás tartalmát és egy-két kivétellel a feladat utasításait sem olvasták el. Az értékelő skála kritériumait viszont a kompetens értékelők többször olvasták fel, mint a profik. Az írásokból majdnem egyenlő mértékben olvastak fel, egy-egy szót viszont a kompetens értékelők olvastak többször.

A nyelvhelyesség pontozását az értékelők alaposan végezték, mind a tizenegy értékelési és egy kivétellel az összes olvasással kapcsolatos stratégiát használták. Egyetlen értékelő sem olvasta a fogalmazás utasítását a nyelvhelyesség értékelésekor. Az értékeléskor használt stratégiák számának eloszlása a tíz íráson jelentősen különbözött, ami arra enged következtetni, hogy a nyelvhelyesség értékelése eltérő figyelmet igényel az írások minősége szerint. Az egyes kritériumokat eltérően alkalmazták a kompetens és a profi értékelők: a kompetensek több figyelmet szenteltek a nyelvtan helyességére, mint a szerkezetek változatosságára, valamint többször olvasták fel a skála kritériumait értékeléskor. A profi értékelők viszont fordítva: kevesebbet olvastak a skálából és inkább a szöveget hasonlították a kritériumhoz. A profi értékelők sokszor fogalmazták meg véleményüket saját szavaikkal. Az értékelők gyakran kiegészítették a meglévő kritériumokat kiegészítették olyanokkal, amiket fontosnak találtak: többen említették a szórendet, utaltak a központozásra, névelők használatára, szóismétlésre és az írás stílusára. A szókincs értékeléséhez hasonlóan, a nyelvhelyesség elbírálásakor is előfordult, hogy főleg a kompetens értékelők nem tudtak különbséget tenni a szókincs és a nyelvhelyesség kritériumai között.

Az értékelő skála negyedik szempontja a szöveg megszerkesztettségének értékelése volt, amelyet a skálán szereplő két kritérium szerint bíráltak az értékelők: mennyiben felel

meg az írás a feladatban előírtaknak, vagyis baráti levélnek, illetve mennyire logikus a levél felépítése. A profi értékelők több figyelmet fordítottak a szöveg logikai felépítésére és a kapcsolószavak meglétére, mint a kompetens értékelők. A kompetens értékelők többször olvasták fel a skála kritériumait mielőtt kiválasztották a pontszámot, mint a profik, ezen kívül, az értékelési folyamat elején többet beszéltek, mint a végén. A profi értékelők véleményüket gyakrabban fogalmazták meg saját szavaikkal, mint a kompetensek. A szöveg megszerkesztettsége értékelésekor az értékelők ritkán módosították döntésüket, ami arra enged következtetni, hogy a kritériumokat megfelelőnek találták a szempont elbírálásához. A szempont megítéléséhez nem használták a hibák azonosításának stratégiáját sem. Amennyiben saját kritériumot adtak a meglévőkhez, az valamelyik kritérium egy-egy részletére vonatkozott, például aláírás megléte a levél végén. Összegezve a szöveg szerkesztettségének értékelését, a leggyakrabban az értékelők értékelési stratégiákat használtak, nagyon kevés olvasással kapcsolatos megjegyzés hangzott el és szinte egyáltalán nem nézték meg a feladat szövegét.

A negyedik kutatási kérdés arra vonatkozott, hogyan értelmezték az értékelők az írásokat. A kérdés megválaszolásához kiválasztottam a leggyengébb (N2) és a legjobb (N6) írást és megvizsgáltam hogyan értékelték őket a kompetens és a profi értékelők.

A tíz írás közül a referenciapontok szerint leggyengébb írást a profi értékelők többsége utolsónak sorolta, a kompetens értékelők véleménye ettől némileg eltért: ők inkább egy másikat soroltak utolsónak. A két csoport között megfigyeltem másfajta véleménykülönbséget is az értékelési folyamatok vizsgálata során. Az értékelés előtti szakaszban inkább a profi értékelők mondtak véleményt az írásról, ugyanakkor megjegyzéseik hasonlóak voltak a kompetens értékelők meglátásaihoz. Az is előfordult ebben az értékelési szakaszban, hogy az egyes értékelők teljesen ellentétes véleményüknek adtak hangot az írás minőségével kapcsolatban: volt aki jónak tartotta, míg más értékelő nem tartotta megfelelőnek. A feladat teljesítését a legtöbb kompetens értékelő a tartalmi pontok vizsgálatával végezte, a kommunikatív cél megvalósítására kevés figyelmet fordítottak, ez arra enged következtetni, hogy könnyebbnek találták a tartalmi pontok számolását, illetve csak arra figyeltek, hogy az író reagált-e a feladat szerinti szempontokra, de hogy hogyan, arra már nem tértek ki. Ezen kívül, ők többször olvastak az írásokból, mint a profi értékelők az értékelési folyamat során. A szókincs értékelése teljesen más mintát eredményezett: a kompetens értékelők kevesebb értékelési stratégiát alkalmaztak, mint a profik, és a profi értékelők többször olvastak az írásból,

mint a kompetens értékelők. Ezek szerint a szókincs értékelése nehéznek bizonyult, amit az is alátámaszt, hogy ugyanazt a szót például az egyik értékelő a széles választékhoz sorolta, míg egy másik a megfelelő kategóriába sorolta. A nyelvhelyesség elbírálásakor a kompetens értékelők pontjai közelebb voltak a referenciaponthoz, míg a profi értékelők magasabb pontszámot választottak. A szöveg megszerkesztettségének megítélése nemcsak a két értékelő csoport között, de csoporton belül is jelentős eltéréseket eredményezett: az értékelők, főleg a kompetensek szélesebb skálán választottak pontszámot, előfordult 0 és öt pont is, míg a profi értékelők is viszonylag változatos pontszámot ítéltek oda (1 és 4 között). Az értékelők gondolatmenetének vizsgálata során megfigyeltem, hogy bár a stratégiák és az érvek hasonlóak voltak, a pontszámok mégis eltértek.

A referenciapontok alapján legjobbnak ítélt írást az értékelők többsége szintén az első helyre sorolta, a profi értékelők pontszámai magasabbak voltak, így közelebb álltak a referenciapontokhoz. Az értékelés előtti szakaszban a profi értékelők több megjegyzést tettek, mint a kompetens értékelők. Az értékelési folyamatokat főleg értékelési és olvasási stratégiák jellemezték, az értékelők ritkán fordultak saját kritériumok használatához. Az értékelők a feladat teljesítését vagy öt vagy hat pontra értékelték, a kompetens értékelők többször olvastak az írásokból mint a profik. A szókincs megítélése elért, a pontszámot a kompetens értékelők szélesebb skáláról választották és több értékelési stratégiát alkalmaztak, a profi értékelők közel egyenlő arányban használtak értékelő és olvasással kapcsolatos stratégiákat. A nyelvhelyesség értékelése is különbözött: a kompetens értékelők többféle pontszámot választottak. A nyelvhelyességet főleg értékelési stratégiák segítségével pontozták, a profi értékelők gyakran fogalmazták meg véleményüket saját szavaikkal. A szöveg megszerkesztettségére a kompetens értékelők magasabb pontszámot adtak és ők inkább a szöveg elrendezésére figyeltek és nem a logikai kapcsolatokra, mint a profi értékelők.

Összefoglalva megállapítható, hogy az értékelők a legjobb írást könnyebben azonosították, mint a leggyengébbet és a legjobb írás értékelési folyamatai több hasonlóságot mutattak a kompetens és a profi értékelők között, mint amikor a leggyengébb írásokat értékelték.

Az utolsó kutatási kérdés az értékelők véleményének elemzését tűzte ki célul. Az értékelők visszajelzéseit három kérdéssel irányítottam: arra kértem őket, véleményezzék a tudásszint mérésével foglalkozó szemináriumot, az íráskészség értékelésére felkészítő egyeztető tréninget és a fogalmazások értékelését. A 37 hallgató három kivételével visszajuttatta a kérdőívet és a válaszadók közül nem mindenki írta le véleményét

mindegyik kérdésre. A visszajelzések között általános és konkrét megjegyzések is találhatók. A hallgatók összességében elégedettek voltak a szemináriummal, értékelték az elmélet és a gyakorlat megfelelő arányát, többen említettek egy-egy konkrét példát is, főleg a profi értékelők. A visszajelzésben többen megjegyezték, hogy a kurzus segített nekik jobban megérteni a különbséget a nyelvtanuló és a nyelvtanár között, valamint sikerült elsajátítaniuk a nyelvtudásmérés alapelveit. A fogalmazások értékelésére felkészítő egyeztető tréningről megoszlottak a vélemények: egyesek elegendőnek találták a gyakorlást, míg mások, főleg a kompetens értékelők szívesen gyakoroltak volna többet. A véleményekben az is tükröződik, hogy a hallgatók több szempontból is hasznosnak találták az értékelő feladatot: megítélésük szerint a tanárképzésben szükség van ilyen jellegű tanulmányokra és egyénileg is sokat hasznosíthattak a hangosan gondolkodással járó feladattól. Sokan említették, hogy a szövegek átírata időigényes munka volt, de sokat tanultak saját gondolataik leírásakor.

4. A kutatás relevanciája

Az íráskészség értékelésével kapcsolatos nemzetközi irodalomban számos tanulmány foglalkozik az értékelők jellemzőivel, valamint az értékelés folyamataival. Ezen kutatások főleg a verbális protokoll módszerét használták az adatok gyűjtésére. Eredményeim igazolják, hogy az íráskészség értékelése olyan problémamegoldó folyamat, amely során az értékelő döntéseket hoz az írás minőségével kapcsolatban és nem egyszerűen összeveti az írást az értékelő skálával (DeRemer, 1998). Az értékelő döntéshozatalát többen is vizsgálták és kialakították az értékelési folyamatok modelljeit.

Milanovic et al. (1996) modellje négy fő szakaszból áll: értékelés előtti, az írás áttekintése, gyors olvasás és értékelés. Ezen szakaszokon belül az értékelők figyelme különböző mértékben érvényesül az írás jellemzőivel kapcsolatban. Kutatásom során bebizonyosodott, hogy az értékelési folyamat szakaszokból épül fel, azonban nem minden esetben figyeltem meg az értékelés előtti szakaszt, azonban az, hogy az értékelők eltérő mértékben figyelnek a szövegek jellemzőire érvényes volt mind a kompetens, mind a profi értékelőkre.

Wolfe (1997) az értékelési folyamat megfigyeléséhez az értékelőket három csoportba osztotta értékelési teljesítményük alapján. Ezt a módszert választottam és a 37 értékelő pontjait a referenciapontokhoz hasonlítottam, így két csoportot alakítottam ki: kompetens értékelőket, akik döntései kevésbé egyeztek a referenciapontokkal, és profikat, akik

pontjai közelebb voltak a referenciapontokhoz. Wolfe modellje szerint az értékelőkben kialakul az írás és az értékelés saját értelmezése és az értékelést ezek alapján végzik, tehát az eredmény függ az értékelők tulajdonságaitól. Wolfe azt vizsgálta, hogy az értékelési folyamatokban megfigyelhető váltások mennyire függenek az értékelők teljesítményétől, melyik értékelő csoport hoz korai döntéseket és kik tesznek több személyes jellegű megjegyzést. Hasonlóan Wolfe eredményeihez megállapítottam, hogy a profi értékelők kevesebbszer váltanak az értékelési kritériumok között. Az azonban, hogy a profi értékelők korábban hoznak döntést, nem igazolódott be, ennek oka valószínű az értékelési skála jellege lehet, ugyanis Wolfe értékelői holisztikus, míg kutatásom résztvevői analitikus skálát használtak. A kompetens értékelők többször tesznek személyes megjegyzést és az írások felszíni tulajdonságaira figyelnek jobban. Eredményeim szerint, a profi értékelők fordultak többször egyéni szempontokhoz, azonban az beigazolódott, hogy a kompetens értékelők inkább a felszíni tulajdonságokra figyelnek.

A legrészletesebb modell Lumley (2000; 2002) kutatásaiban található, vizsgálatában tapasztalt értékelőket figyelt meg. Megállapította, hogy az értékelési folyamatok három egymástól jól elkülönülő szakaszból állnak: első olvasás, értékelés a négy szempont szerint és a pont odaítélése. Az értékelési folyamatot ezenkívül három szinten képzelem el: az értékelés középpontjában az instrumentális szint található, amely magában foglalja az értékelést az összes összetevővel. Az intézményi szint az értékelés szociális vetülete, vagyis azok a folyamatok, amelyek az értékelés kontextusából adódnak. A harmadik, értelmező szint azokat a magatartásformákat tartalmazza, amiket az értékelők konfliktus esetén alkalmaznak, vagyis akkor, ha az írás és a kritériumok között konfliktus keletkezik. Lumley értékelőihez hasonló stratégiákat figyeltem meg a profi és kompetens értékelők folyamataiban, valamint azt, hogy az értékelés középpontjában az instrumentális szint található, és az értékelési folyamatokat ismétlődő lépések jellemzik.

Cumming et al. (2002) modellje három egymást követő tanulmány eredményeinek alapján jött létre. A modell az értékelési stratégiákat két csoportba osztja: értelmező és értékelő stratégiákra. Az értelmezés az írás olvasását és értelmezését foglalja magába, az értékelés pedig azokat a stratégiákat, amelyek segítségével az értékelők meghozzák döntéseiket. Kutatásomban a stratégiákat a négy értékelési szempont és az értékelők fókusza szerint csoportosítottam és ezek között én is megfigyeltem értelmezési és értékelési stratégiákat.

A fenti modellek különböző kutatások eredményeként születtek, különböző értékelőkkel, írásokkal, skálákkal és értékelési helyzetekben. Kutatásom eredményei

számos esetben mutattak hasonlóságot a korábbiakkal, ami azt is igazolja, hogy az íráskészség értékelése bonyolult kognitív feladat, azonban, főleg a verbális protokoll módszerével jól megfigyelhető és az értékelők gondolkodása nyomon követhető.

Magyarországon az idegen nyelvi tudásszint mérésének irodalmában kevés az íráskészség értékelésével, ezen belül az értékelők viselkedésével foglalkozó tanulmány. Az adatgyűjtés verbális protokoll segítségével is újszerű hazánkban, legutóbb Nikolov (2006) elemezte diákok feladatmegoldását hangosan gondolkodás módszerével.

Jelen disszertáció relevanciája tehát, egyrészt az íráskészséget értékelők gondolkodásának megfigyelésében rejlik. Másrészt, az adatgyűjtés verbális protokoll segítségével történt, ami kevésbé használt módszer az idegen nyelvi tudásszint mérésében hazánkban. Végül, a kutatásban résztvevő értékelők száma 37, ami mind a nemzetközi, mind a hazai szakirodalomban ritka a hasonló jellegű kutatásokban.

5. A kutatás korlátai

Annak ellenére, hogy 37 résztvevő értékelési folyamatait elemeztem, a kutatás eredményeiből levont következtetések főleg a vizsgált adatokra vonatkoznak, a kutatás jellege csak óvatos általánosításokra adhat alkalmat, azonban jól megfigyelhető tendenciák bontakoztak ki, amelyek hasznosíthatók esetleges további kutatásokban. Az elemzéseket az értékelők verbális protokolljai és a pontszámai vizsgálatával végeztem. A verbális protokoll, mint kutatási módszer alkalmas gondolkodási műveletek megfigyelésére, azonban nem minden gondolatot tudnak az emberek hangosan kimondani (Ericsson & Simon, 1993). Ezen kívül, az értékelt írások száma tíz volt összesen, ami a gondos válogatás ellenére sem biztos, hogy hűen reprezentálja a vizsgált íráskészséget. Ez annak ellenére, hogy a kutatás középpontjában az értékelési folyamatok álltak, nem az íráskészség, befolyásolhatta az eredményeket. A hasonló kutatások (Cumming et al., 2002; Lumley, 2000; Milanovic et al., 1996; Wolfe, 1997) résztvevői általában gyakorlott vizsgáztatók és a kutatás vizsgahelyzetéhez közeli kontextusban zajlik. Jelen kutatásban tanítási és vizsgáztatási tapasztalattal nem rendelkező értékelők vettek részt, ez is azt támasztja alá, hogy a következtetések főleg erre a kutatásra érvényesek.

6. A téma további kutatásának lehetőségei

Az összegyűjtött adatokat további szempontok szerint is lehetne elemezni. Például az egyes értékelők esettanulmányai és azok összehasonlítása mélyebb betekintésre adna lehetőséget az értékelési folyamatokba. A kutatást ki lehetne egészíteni az értékelők kikérdezésével is, amiben megindokolnák mit és miért mondtak.

Az íráskészség vizsgálatában számos további irány is elképzelhető. A kutatás tanári és vizsgáztatói tapasztalattal nem rendelkező értékelőket vizsgált, ezért elképzelhető tapasztalt értékelők vizsgálata „éles” vizsgahelyzetben, főleg a magyarországi angol érettségi vizsga kutatása járulna hozzá jelentősen a hazai nyelvtanítás fejlesztéséhez.

7. Az értekezés témájában megjelent publikációk

- Bukta, K. (2000). Reflections on test-taking strategies of 7th and 11th grade Hungarian students of English. *novELTy*, 7(3), 48-59.
- Bukta, K. (2001). Mit tanulnak a diákok angol órán? 7. és 11. évfolyamos tanulók angol nyelvtudásának vizsgálata. *Iskolakultúra*, 11(8), 36-48.
- Bukta, K. (2007). Assessment of written performance: Tracing raters' decision making process. *Porta Linguarum*, 8(2), 21-41.
- Bukta, K., & Nikolov, M. (2002). Nyelvtanítás és hasznos nyelvtudás: Az angol mint idegen nyelv. In Csapó, B. (Ed.), *Az iskolai műveltség* (pp. 169-192). Budapest: Osiris.
- Bukta, K., Gróf, Sz., & Sulyok, A. (2005). *7 próba érettségi angol nyelvből. Középszint*. Szeged: Maxim Kiadó.

8. Hivatkozott irodalom

- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Common European framework of reference for languages: Learning, teaching, assessment*. (2001). Cambridge: Cambridge University Press.
- Cumming, A., Kantor, R., & Powers, D. E. (2002). Decision making while rating ESL/EFL writing tasks: A descriptive framework. *The Modern Language Journal*, 86(1), 66-96.
- DeRemer, M., L. (1998). Writing assessment: raters' elaboration of the rating task. *Assessing writing*, 5(1), 7-29.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis. Verbal reports as data*. (Rev. ed.). Cambridge, MA: MIT Press.
- Fekete, H., Major, É., & Nikolov, M. (Eds.) (1999). *English language education in Hungary. A baseline study*. Budapest: British Council.

- Gass, S., & Mackey, A. (2007). *Data elicitation for second and foreign language research*. London: Lawrence Erlbaum Associates, Publishers.
- Imre, A. (1995). Tendenciák az elmúlt évek nyelvoktatásában. *Iskolakultúra* 5(15-17), 146-153.
- Katona, L. (1998). Criteria for the design and evaluation of communicative language tests. *novELTy*, 5(4), 47-53.
- Kiszely, Z. (2003). *Students' writing in L1 Hungarian and L2 English: rhetorical patterns, writing processes and literacy backgrounds*. Unpublished PhD dissertation, Pécs.
- Kontráné, H. E. (2003). Training teachers to mark compositions. *novELTy*, 10(3), 5-24.
- Lumley, T. (2000). *The process of the assessment of writing performance: the rater's perspective*. Unpublished PhD dissertation. Department of Linguistics and Applied Linguistics: The University of Melbourne.
- McNamara, T. F. (1996). *Measuring second language performance*. London and New York: Longman.
- Medgyes, P. (1992). Angol – a kommunikáció pótnyelve. Körkép az angol nyelv magyarországi oktatásáról és terjedéséről. *Magyar Pedagógia* 92(4), 263-283.
- Milanovic, M., Saville, N., & Shuhong, S. (1996). A study of decision-making behaviour of composition markers. In Milanovic, M., & Saville, N. (Eds.). *Studies in Language Testing 3: Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arhem*. (pp. 92-115). Cambridge: Cambridge University Press.
- Nikolov, M. (2006). Why whales have migraine. *Language Learning*, 56(1), 1-51.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wolfe, E. W. (1997). The relationship between essay reading style and scoring proficiency in a psychometric scoring system. *Assessing Writing*, 4(1), 83-106.